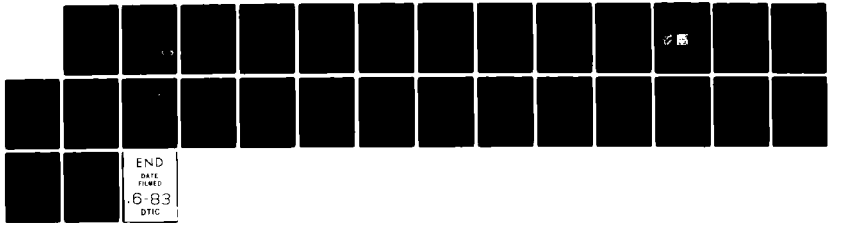
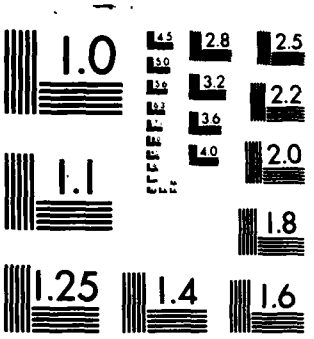
AD-A128 398 THE MEASUREMENT OF VISUAL MOTION(U) MASSACHUSETTS INST 1/1
OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB
E C HILDRETH ET AL. DEC 82 AI-M-699 N00014-80-C-0505
UNCLASSIFIED F/G 6/4 NL



END
DATE FILMED
6-83
DTIC



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>A.I. Memo No. 699 | 2. GOVT ACCESSION NO.<br>AD-A128398 | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>The Measurement of Visual Motion | | 5. TYPE OF REPORT & PERIOD COVERED<br>memorandum<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Ellen C. Hildreth and Shimon Ullman | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-80-C-0505,<br>79-23110 MCS |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Artificial Intelligence Laboratory<br>545 Technology Square<br>Cambridge, Massachusetts 02139 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Advanced Research Projects Agency<br>1400 Wilson Blvd<br>Arlington, Virginia 22209 | | 12. REPORT DATE<br>December 1982<br>13. NUMBER OF PAGES<br>26 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)<br>Office of Naval Research<br>Information Systems<br>Arlington, Virginia 22217 | | 15. SECURITY CLASS. (of this report)<br>UNCLASSIFIED<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

motion measurement
velocity field
optical flow
zero-crossings
motion perception

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The analysis of visual motion divides naturally into two stages: the first is the measurement of motion, for example the assignment of direction and magnitude of velocity to elements in the image, on the basis of the changing intensity pattern; the second is the use of motion measurements, for example, to separate the scene into distinct objects, and infer their three-dimensional structure. In this paper, we present a computational study of the measurement of motion. Similar to other (over)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Dat

Block 20. (continued)

visual processes, the motion of elements is not determined uniquely by information in the changing image; additional constraint is required to compute a unique velocity field. Given this global ambiguity of motion, local measurements from the changing image, such as those provided by directionally-selective simple cells in primate visual cortex, cannot possibly specify a unique local velocity vector, and in fact, specify only one component of velocity. Computation of the full two-dimensional velocity field requires the integration of local motion measurements, either over an area, or along contours in the image. We will examine possible algorithms for computing motion, based on a range of additional constraints. Finally, we will present implications for the biological computation of motion.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 699 December, 1982

# THE MEASUREMENT OF VISUAL MOTION

**Ellen C. Hildreth and Shimon Ullman**

ABSTRACT: The analysis of visual motion divides naturally into two stages: the first is the measurement of motion, for example, the assignment of direction and magnitude of velocity to elements in the image, on the basis of the changing intensity pattern; the second is the use of motion measurements, for example, to separate the scene into distinct objects, and infer their three-dimensional structure. In this paper, we present a computational study of the measurement of motion. Similar to other visual processes, the motion of elements is *not* determined uniquely by information in the changing image; additional constraint is required to compute a unique velocity field. Given this global ambiguity of motion, local measurements from the changing image, such as those provided by directionally-selective simple cells in primate visual cortex, cannot possibly specify a unique local velocity vector, and in fact, specify only one component of velocity. Computation of the full two-dimensional velocity field requires the integration of local motion measurements, either over an area, or along contours in the image. We will examine possible algorithms for computing motion, based on a range of additional constraints. Finally, we will present implications for the biological computation of motion.



This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-80-C-0505 and in part by National Science Foundation Grant 79-23110MCS.

## 1. Introduction

The organization of movement in a changing two-dimensional image provides a valuable source of information for analyzing the environment in terms of objects, their motion in space, and their three-dimensional structure. It is not surprizing, therefore, that the analysis of visual motion plays a central role in biological vision systems. Sophisticated mechanisms for extracting and utilizing motion exist even in simple animals. For example, the frog has efficient "bug detection" mechanisms that respond selectively to small, dark objects moving in its visual field [1]. The ordinary housefly can track moving objects and discover the relative motion between a target and its background, even when the two are identical in texture, and therefore indistinguishable in the absence of relative motion [2].

In higher animals, including primates, the analysis of motion is "wired into" the visual system from the earliest processing stages. Some species, such as the pigeon [3] and rabbit [4] (see [5] for other examples) perform rudimentary motion analysis at the retinal level. In other animals, including cats and primates, the first neurons in visual cortex to receive input from the eyes are already involved in the analysis of motion: they respond well to stimuli moving in one direction, but little, or not at all, to motion in the opposite direction [6,7].

In some animals, visual motion is used in the guidance of locomotion and the control of body motion. The plummeting gannet [8], for example, uses visual flow information to stretch back its wings a fraction of a second before it hits the water. Perhaps the most remarkable use of visual motion is the recovery of three-dimensional shape using motion information alone. This capacity of the human visual system has been demonstrated in the studies of Wallach and O'Connell [9] and Johansson [10,11].

The extensive use of motion by biological systems, and in particular the human visual system, demonstrates the feasibility of carrying out certain information processing tasks and helps to establish specific goals for the analysis of time-varying imagery. This analysis divides naturally into two parts. The first stage is the measurement of motion; for example, the assignment of direction and magnitude of velocity to elements in the image, on the basis of the changing intensity pattern. The second is the use of motion measurements; for example, to separate the scene into distinct objects, and infer their three-dimensional structure.

In this paper, we present a computational study of the measurement of visual motion. It is a problem which was found to be surprizingly difficult, both in computer vision, and in modelling biological vision systems. We will present the general problem of motion measurement in Section 2, and discuss methods that have been proposed for its solution. Section 3 presents a specific scheme, proposed by Marr and Ullman [12], for extracting the first motion measurements from the changing image. The initial measurements do not yet specify the true motion of objects in the changing image, and must be combined in some way. This raises the *motion integration problem*, which will be discussed in Sections 3 and 4. Section 5 presents some implications for the analysis of motion in biological vision systems.

## 2. Motion Detection and Measurement

The motion of elements and regions in an image is not given directly, but must be computed from more elementary measurements. The initial registration of light by the eye or by electronic imaging devices can be described as producing a

two-dimensional array of time-dependent light intensity values, $I(x, y, t)$. Motion in the image can be described in terms of a vector field $\mathrm{V}(x, y, t)$ giving the velocity of a point with image coordinates $(x, y)$ at time $t$. The first problem in analyzing visual motion is the computation of $\mathrm{V}(x, y, t)$ from $I(x, y, t)$. This computation is the measurement of visual motion.

In some cases it may be sufficient to detect only certain properties of the velocity field, rather than measure it completely and precisely. For example, in order to respond quickly to a moving object, motion must be detected, but not necessarily measured. Other tasks, such as the recovery of three-dimensional structure from motion, require a more complete and accurate measurement of the velocity field [13-17].

The measurement of motion may be performed at different stages in the processing of an image, utilizing different motion primitives. It is useful to draw a distinction between two main schemes. At the lowest level, motion measurements may be based directly on the local changes in light intensity values; these are called *intensity-based* schemes. Alternatively, it is possible to first identify features such as edges and their termination points, corners, blobs, or regions, and then measure motion by matching these features over time, and detecting their changing positions. Schemes of this type are called *token-matching* schemes. These two modes of motion detection and measurement give rise to different computational problems, and consequently to different kinds of processes in biological as well as computer vision systems.

### 2.1. Intensity-based Schemes for Motion Measurement

Two main types of intensity-based schemes have been advanced for biological and computer vision systems: correlation techniques and gradient methods. Cross-correlation of raw intensity values has been used in computer vision applications [18-21], and has been proposed as a model for motion measurement in the human visual system [22-24]. Related to cross-correlation schemes are subtraction schemes, involving simple differencing operations between successive frames. In computer vision, such schemes are primarily used for the detection of motion, and object segmentation [25-28]; together with cross-correlation, they have been utilized for the measurement of motion [26,28]. A fundamental problem of most correlation and subtraction schemes is that they assume the image (or a large portion of it) moves as a whole between the two frames. Images containing independently moving objects and image distortions induced by the unrestricted motion of objects in space pose difficult, perhaps insurmountable, problems for these techniques.

Other intensity-based schemes have been proposed for biological systems. A simple motion detector can be constructed by comparing the outputs of two detectors to light increments at two adjacent positions. The output at position $p_1$ and time $t$ is compared with that at $p_2$ at time $t - \delta t$ (a low-pass temporal filter may be used instead of the delay [29]). Two variations of this approach, called the *delayed comparison* scheme, have been proposed as models for biological systems. The first is obtained by multiplying the two values, i.e. $D(p_1, t) \cdot D(p_2, t - \delta t)$, where $D$ denotes the output of the subunits, shown in Figure 1a. If a point of light moves from $p_2$ to $p_1$ in time equal to $\delta t$, this product will be positive. In an array of such detectors, the average output is essentially equivalent to a cross-correlation of the inputs [29]. An alternative method along the same general line is the "And-Not" scheme proposed by Barlow and Levick [30] for the directionally selective units in the rabbit's retina (a similar scheme was suggested for the cat's visual cortex [31]).

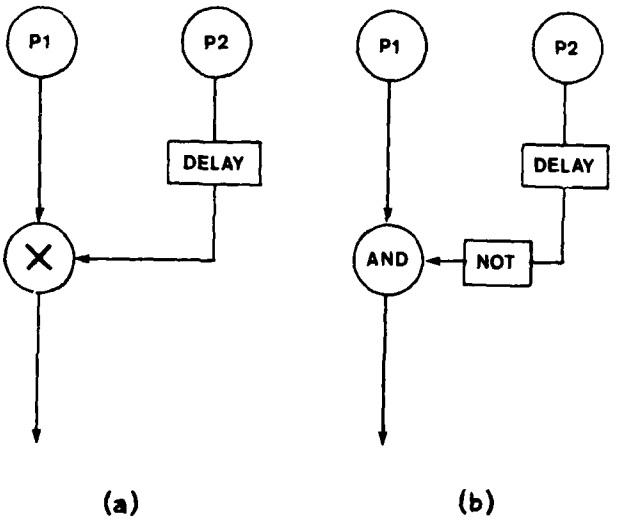



**Figure 1.** The delayed comparison schemes. (a) The two inputs are multiplied (b) The "And-Not" scheme

Evidence for inhibitory interactions within the directionally selective mechanism led to a model in which the motion detector computes the logical "And" of $D(p_1, t)$ and "Not" of $D(p_2, t - \delta t)$ (see Figure 1b). In this scheme, a motion from $p_2$ to $p_1$ is "vetoed" by a delayed response from $p_2$, whereas motion from $p_1$ to $p_2$ produces a positive response. Poggio and Reichardt [32] have proposed a similar scheme for the visual system of the fly, and an elegant synaptic mechanism that implements these computations was described by Torre and Poggio [33].

Some general properties of the delayed comparison schemes are worth noting. First, these detectors respond selectively not only to continuous motion, but also to discrete jumps of the stimulus between positions $p_1$ and $p_2$. Second, the speed of motion must lie within a certain range, determined by the delay (or the low-pass filtering) and the separation between the receptors. A range of velocities can be covered by a family of detectors with different internal delays and interreceptor spacing. Finally, motion measurements cannot be determined reliably from the output of a single detector of this type. The accurate and reliable measurement of motion will require the combination of the outputs from an array of such elementary detectors.

In gradient schemes, the local motion measurements are derived via a comparison between intensity gradients, and temporal intensity changes. A one-dimensional example, illustrating the basic principle, is shown in Figure 2. Consider the intensity profile (intensity $I$ as a function of position $x$), indicated by the solid curve in Figure 2. At the point $p$, the profile has a positive slope. If the profile moves to the left, indicated by the dashed curve, the intensity value $I$ at $p$ will be increasing; for a rightward motion, indicated by the dotted and dashed curve, $I(p)$ will be decreasing. The sign of the temporal change in $I(p)$ thus signals the direction of motion, and from the magnitude of the spatial and temporal intensity changes, the speed of motion can be determined. In principle, measurements of motion may be obtained wherever the image intensity gradient is non-zero; however, the measurements are more reliable at the location of edges, where steep intensity gradients are induced.

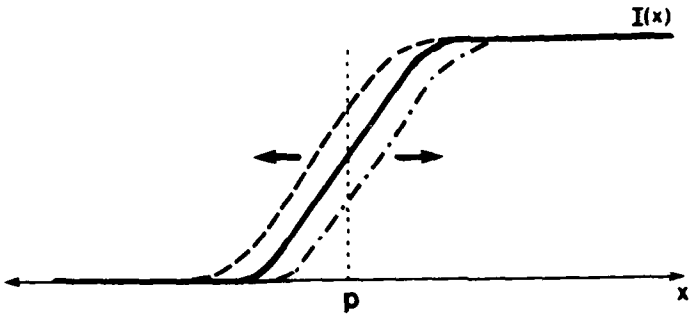



**Figure 2.** Comparison of the sign of the spatial and temporal derivatives of intensity at the point *p* yields the sign of direction of motion

In two dimensions, the spatial and temporal intensity changes alone are not sufficient to determine the local direction and magnitude of velocity [12, 34-37], because of the *aperture problem*, illustrated in Figure 3. If the motion of the edge E is to be detected by operations which examine an area A that is small compared to the overall extent of the edge, the only motion that can be extracted is the component **c** perpendicular to the local orientation of the edge. For example, such operations cannot distinguish between motion in the directions **b**, **c**, and **d**. To determine the motion completely, a second stage of analysis is required, which integrates the local motion measurements, either over an area of the image, or along contours.

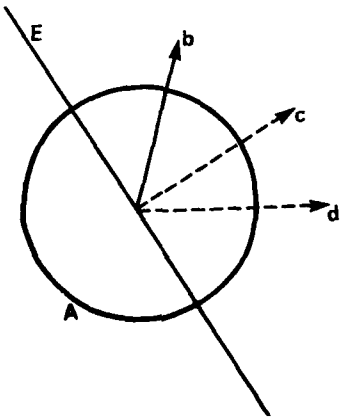



**Figure 3.** The aperture problem. Motion in the directions **b**, **c** and **d** can not be distinguished when viewed through the local aperture **A**.

## 2.2. Token-matching Schemes for Motion Measurement

In token-matching schemes, identifiable elements – tokens – are located and then matched over time. Assuming that the visual input is given as a sequence of discrete frames, a counterpart for each element in one frame must be located in the next. This raises the *correspondence problem*, illustrated in Figure 4. The

filled circles in the figure represent the first frame, and the open circles the second. There are two possible one-to-one pairings between the elements of the two frames, leading to two patterns of perceived motion: diagonal (a) or horizontal (b). In this example, the match is only two-way ambiguous. In general, each frame could contain many elements arranged in complex figures; a correspondence must then be established among them. The rules governing the correspondence process in human vision have been investigated [38-44], but are still far from being completely understood. Token-matching schemes for motion measurement have also been studied for computer vision [45-50].

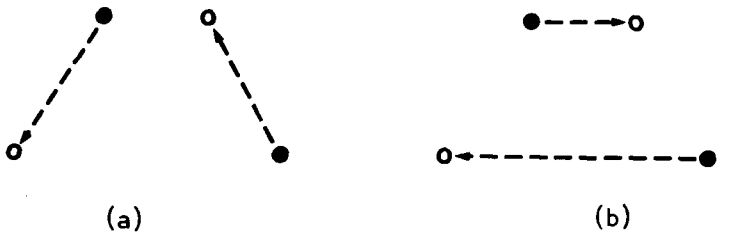



**Figure 4.** A simple correspondence problem

Two general problems of token-matching schemes are relevant to both biological and machine motion analysis. The first concerns the level at which the correspondence is established. By this we mean the degree of preprocessing and the complexity of the participating tokens. Matching may be established between simple tokens such as points, blobs, and edge fragments. Alternatively, the matching process may operate on complex tokens such as structured forms, or even the images of recognized objects. The use of complex tokens can simplify the correspondence process, since a complex token will usually have a unique counterpart in a subsequent frame. Primitive tokens will usually have many competing possible matches, but their use has two distinct advantages. The first is a reduced preprocessing requirement, which is of special importance in motion perception, where computation time is severely restricted. The second is that a correspondence scheme based on primitive tokens can operate on arbitrary objects engaged in complex shape changes. It seems, therefore, that the correspondence process should operate on the level of rather primitive elements, perhaps at the level of Marr's full primal sketch [51,52].

The second general problem concerns the possible role of intensity-based and token-matching schemes in an integrated vision system. Intensity-based schemes tend to be fast and sensitive, but the ambiguity of the local measurements may make it difficult to recover the velocity field accurately. A token-matching scheme can, in principle, track a sharply localized token (such as a line termination) over long distances, and thereby achieve a high degree of accuracy, at the price of more extensive processing, in locating the tokens and solving the correspondence problem.

In light of the differences in their basic properties, it is possible that the two motion measurement schemes serve distinct visual tasks. The intensity-based system may be useful as an "early warning" system, and for the separation of moving objects from their background. Token-matching schemes may play an important role in the recovery of structure from motion, where the accurate tracking over considerable distances is useful. A second possiblility is that the two schemes interact to complement each other. For example, a token-matching scheme might be guided by additional constraints supplied by an intensity-based system.

## 2.3. Two Motion Systems in Human Vision

Psychological studies of motion detection and measurement in the human system have distinguished two types of visual motion: discrete and continuous. For human observers to perceive motion, the stimulus need not move continuously across the visual field. Under the appropriate spatial and temporal presentation parameters, a stimulus presented sequentially can produce the impression of smooth, uninterrupted motion (as in motion pictures) [53]. The visual system can "fill-in" the gaps in the discrete presentation even when the stimuli are separated by up to several degrees of visual angle, and by long temporal intervals (400 msec., [54]). The resulting motion, termed "apparent" or "beta" motion is perceptually indistinguishable from continuous motion.

The apparent motion phenomena raise the question of whether discrete and continuous motion are registered by two different mechanisms. The fact that the visual system can register both types of motion does not imply the existence of two separate mechanisms, since a system that registers discrete motion could in principle register continuous motion as well. Psychophysical evidence supports, however, the view that two different mechanisms are in fact involved in the process of motion detection and measurement [55-60]. The terms "short range" and "long range" processes were suggested by Braddick [58] for the two mechanisms. The short range mechanism registers continuous motion, or motion presented discretely, with displacements of up to about 15 min. of arc (in the center of the visual field) and temporal intervals of less than about 60–100 msec. The long range mechanism can process larger displacements and temporal intervals. Braddick's terminology characterizes the distinction between the two mechanisms better than the discrete/continuous dichotomy, since discrete presentation with jumps of up to 15 min. of visual arc will be processed by the short range mechanism.

In the human visual system, it appears that the short range process is an intensity-based scheme, whereas the long range process is a token-matching scheme. Braddick [58] proposed that the directionally-selective units of visual cortex underly the short range process, suggesting that the spatial and temporal limits reflect the spatial and temporal parameters of these neural units. Marr and Ullman [12], present a gradient scheme for the detection and measurement of motion, which includes a model for constructing the directionally-selective units, and an algorithm for combining the local measurements to compute the two-dimensional velocity field. The long range motion phenomena illustrate our ability to derive a correspondence of elements in the changing image, over considerable distances and temporal intervals. In these situations, there is no continuous motion of elements across the retina to be measured directly. Psychophysical studies have shown the long range correspondence to be based on more symbolic primitives, such as edges, bars, blobs, simple groups of primitive elements, and texture edges [13,61].

## 2.4. Summary

To summarize, several methods are available for the detection and measurement of motion. These methods differ in the constraints they derive from the changing image. Intensity-based schemes utilize the spatial and temporal changes in the image intensity pattern to constrain local velocity, while token-matching schemes extract more symbolic tokens from the image, which are then matched over time. These two techniques for motion analysis give rise to different computational problems, and consequently to different kinds of processes in biological and computer vision systems.

## 3. Deriving Velocity Constraints from the Image

In this section, we first present a scheme for extracting initial motion constraints from the image, proposed by Marr and Ullman [12], which was motivated by computational studies of early visual processing, and neurophysiological studies of directionally-selective simple cells in primate visual cortex. The use of this type of initial motion measurement raises the *motion integration problem*: the measurements do not yet specify the true motion of objects in the changing image, and must be integrated in some way to compute the velocity field. Computational studies suggested that the first stage of image analysis should be the detection of intensity changes (see [62] for a review). Marr and Hildreth [63] have proposed that an optimal operator for the initial filtering of the image is the Laplacian of a Gaussian, $\nabla^2 G$, whose shape may be approximated by the difference of two Gaussians. The elements in this convolution output, which correspond to the location of intensity changes, are the zero-crossings [64]. Figure 5 shows an image which has been processed through a $\nabla^2 G$ filter, and the resulting zero-crossing contours. Marr and Hildreth suggested that the convolution of the image with $\nabla^2 G$ is represented in the output of the retinal ganglion X-cells, and that a class of simple cells in visual cortex assumes the role of zero-crossing detection.

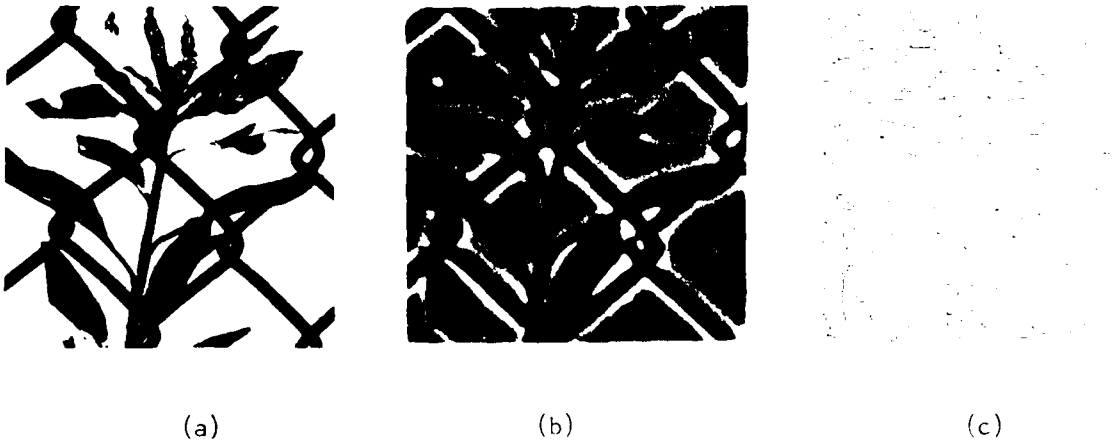



**Figure 5.** The detection of intensity changes. (a) The original image (b) The convolution of (a) with a $\nabla^2 G$ operator (c) The resulting zero-crossing contours.

Marr and Ullman [12] have extended this model for simple cells, including a mechanism for their directional selectivity. The basic idea is illustrated in Figure 6. Figure 6a shows the one-dimensional output of the convolution of a step-edge intensity profile, with the second derivative of a gaussian, $(D^2G^*I)$. Figure 6b and Figure 6c illustrate the time derivative, $\frac{\partial}{\partial t}(D^2G^*I)$, for motion of the profile to the left and right, respectively. At the location of the zero-crossing $Z$, the time derivative will be negative for motion to the left, and positive for motion to the right. Similar to the gradient scheme introduced in Section 2, the sign of contrast of the zero-crossing can be compared with the sign of the temporal derivative, to compute the direction of motion of the zero-crossing. By combining the magnitude of the slope of the convolution output as it crosses zero, with the magnitude of the time derivative, rough magnitude of velocity can be computed. In two dimensions,

comparison of the spatial and temporal derivatives of $\nabla^2 G * I$ (where $I$ is now a two-dimensional intensity distribution) at the location of zero-crossings, provides only the component of motion in the direction perpendicular to the local orientation of the contour.

Marr and Ullman have proposed that the retinal ganglion Y-cells carry the time derivative of the $\nabla^2 G$ convolution, and that simple cells combine the spatial and temporal derivatives, carried by the X-system and Y-system (via the LGN), to compute the direction of motion of the zero-crossing contours. A neural model for the derivation of the spatial and temporal derivatives has been proposed by Richter and Ullman [65]. Recent neurophysiological studies support the role of simple cells in the detection of zero-crossings [Richter, personal communication]. In addition to neurophysiological support, this scheme appears to be consistent with psychophysical studies of the short-range process [12].

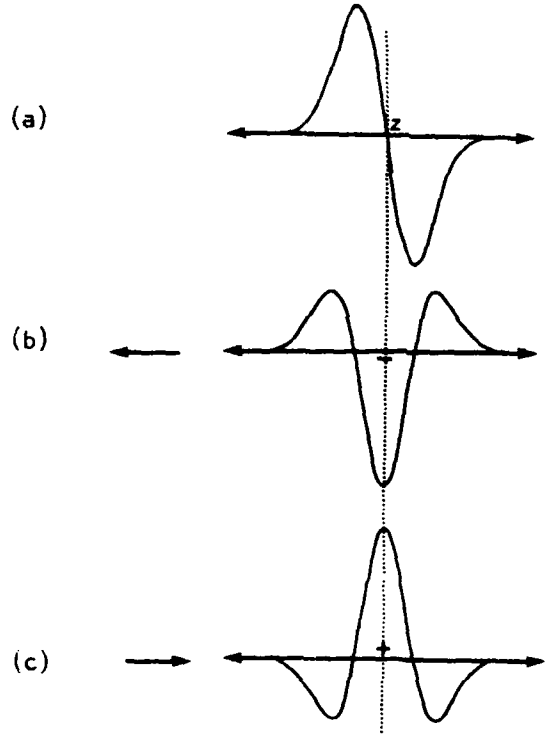



**Figure 6.** The Marr-Ullman scheme. (a) Convolution of a step intensity change with $D^2G$ (b) and (c) Temporal intensity derivative for motion of the profile to the left and right

From a computational standpoint, restricting the measurement of motion to the location of zero-crossings has two advantages over schemes based only on the raw intensities. First, the zero-crossings of $\nabla^2 G * I$ correspond to points in the image at which the gradient of intensity is locally maximum, yielding the most reliable local velocity measurements. Second, the zero-crossings are tied more closely to physical features; if the zero-crossings move, it is more likely to be the consequence of

movement of an underlying physical surface. There are many factors that can cause intensity to change locally, such as changing illumination; a change in intensity over time is not necessarily due to the motion of an underlying surface.

The zero-crossing scheme presented above does not yet solve the motion measurement problem. The measurement of the motion of zero-crossings, using a local gradient scheme, provides only the component of motion in the direction perpendicular to the orientation of the contour. The component of velocity along the contour remains undetected. More formally, we may express the velocity field along a contour by the function $V(s)$, where $s$ denotes arclength. $V(s)$ can be decomposed into components tangent and perpendicular to the contour, as illustrated in Figure 7. $\mathbf{u}^{\top}(s)$ and $\mathbf{u}^{\perp}(s)$ are unit vectors in the directions tangent and perpendicular to the curve, and $v^{\top}(s)$ and $v^{\perp}(s)$ denote the magnitudes of the two components:

$$V(s) = v^{\top}(s)\mathbf{u}^{\top}(s) + v^{\perp}(s)\mathbf{u}^{\perp}(s) \tag{1}$$

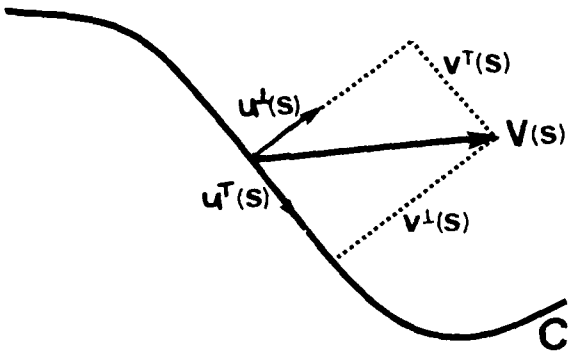
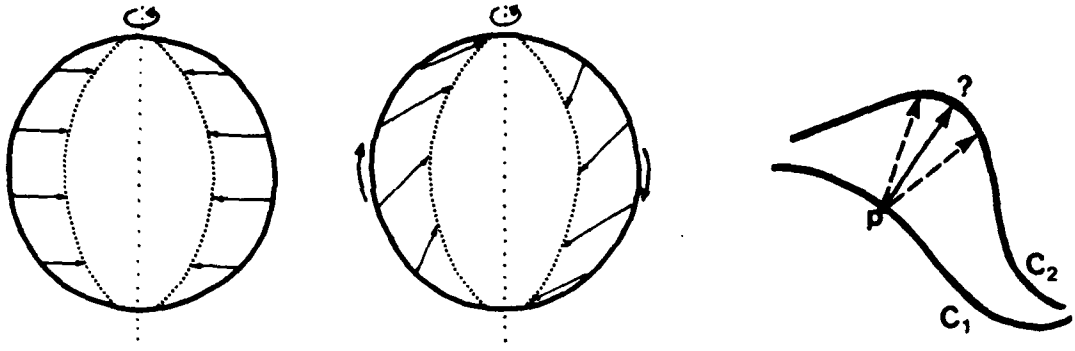



**Figure 7.** The decomposition of velocity $V(s)$ into tangential and perpendicular components

The component $v^{\perp}(s)$ is given directly by the initial measurements from the changing image; the computation of $V(s)$ requires the further recovery of $v^{\top}(s)$.

At the very least, the computation of $V(s)$ requires the integration of the constraints provided by $v^{\perp}(s)$ along the contour. In general however, the solution may still be underdetermined. Additional constraint is required to compute a single velocity field. Figure 8 illustrates two examples, in which the velocity field solution is not unique. In Figure 8a, the solid and dotted lines represent the image of a moving circle, at different instants of time. In the first frame (solid line), the circle lies parallel to the image plane, while in the second frame, the circle is slanted in depth. One velocity field consistent with this sequence is derived from pure rotation of the circle about the central vertical axis, as shown to the left in Figure 8a. (The arrows represent local velocities.) However, there could also be a component of rotation in the plane of the circle, about its center; as shown to the right in Figure 8a. Both velocity fields correspond to valid rigid motions of the circle. This ambiguity is not particular to circles. In Figure 8b, the solid curve $C_1$ rotates, translates and deforms over time, to yield the dotted curve $C_2$. The mapping of points from $C_1$ to $C_2$ is much less clear (consider, for example, different possible velocities for the point $p$). The precise computation of the velocity field in this case is important, when one considers the subsequent computation of structure from




(a) (b)

**Figure 8.** Ambiguity of the velocity field computation. (a) A circle, rotating in depth (b) A deforming curve

motion; different choices for the velocity field may yield different three-dimensional structures. The computation of a unique velocity field requires additional assumptions about physical surfaces, and the velocity field that they generate under motion.

In conclusion, the computation of the velocity field, for the case of general motion, requires a scheme that combines local measurements of motion from the changing image, subject to additional constraints. This is the *motion integration problem.*

## 4. The Integration of Local Motion Measurements

In this section, we discuss the motion integration problem, strictly from a computational viewpoint. The results that we present here are largely independent of the nature of the initial motion measurements, and in particular, do not depend on the Marr-Ullman scheme discussed previously. This section will be organized by the type of additional constraint that may be utilized in the combination stage. We will consider four types of additional constraint on the velocity field: (1) velocity is constant over an area of the image (valid for pure translation); (2) the velocity field is consistent with rigid rotation and translation of objects in the image plane; (3) the velocity field is smooth, and exhibits the least variation among the set of velocity fields consistent with the image constraints; and (4) the velocity field is smooth, exhibits the least variation possible, and is constant over small time intervals. We will discuss methods for combining local measurements, given each of the four types of constraint.

### 4.1. The Constant Velocity Constraint

Much of the previous work in motion analysis has addressed the case of pure translation of objects in the image plane. The early gradient schemes used in computer vision [34,66] assumed that velocity would be constant over a large area of the image. Most correlation and subtraction schemes also embodied this assumption. Marr and Ullman [12,67] proposed a scheme in which each local measurement restricts the true velocity of a patch to lie within a $180^0$ range of directions to one side of a segment of the local zero-crossing contour. A set of measurements taken at different orientations along a zero-crossing contour then

further restrict the allowable velocity directions, until a single velocity direction is obtained, which is consistent with all the local measurements.

The scheme that we present in the next section, for analyzing rotation and translation in the image plane, may also be used for the restricted case of pure translation. While these schemes cannot account for the full range of human motion perception, they may be useful for the initial detection and rough measurement of motion in the periphery, or analysis of motion during smooth pursuit eye movements, in which stationary objects translate rigidly with respect to the eye. In computer vision, there are restricted applications for these techniques, such as the tracking of objects along a conveyor, or computation of camera motion [68].

### 4.2. Rigid Motion in the Image Plane

In this and remaining sections, we will focus on the motion of contours. The results apply, however, to continuous patches in the image as well. First, suppose we have a rigid curve undergoing general motion in space. Its instantaneous motion may be described as the combination of: (1) a rotation with angular velocity $\omega$ about a single axis in space, which we will denote by the vector $\mathbf{n} = [n_1, n_2, n_3]^T$ ($T$ denotes the transpose of a vector), and (2) a translation, which we will denote by the vector $\mathbf{d} = [d_1, d_2, d_3]^T$. Let the curve be given parametrically by $C = (x(s), y(s), z(s))$. The location of a point on the curve may be given by the position vector $\mathbf{r} = [x(s), y(s), z(s)]^T$. If we let the optical axis lie along the $z$-axis, and let the projection of the curve onto the image plane (the $(x, y)$ plane) be orthographic, then the two-dimensional velocity field $\mathbf{V}(s)$ along the contour is given by:

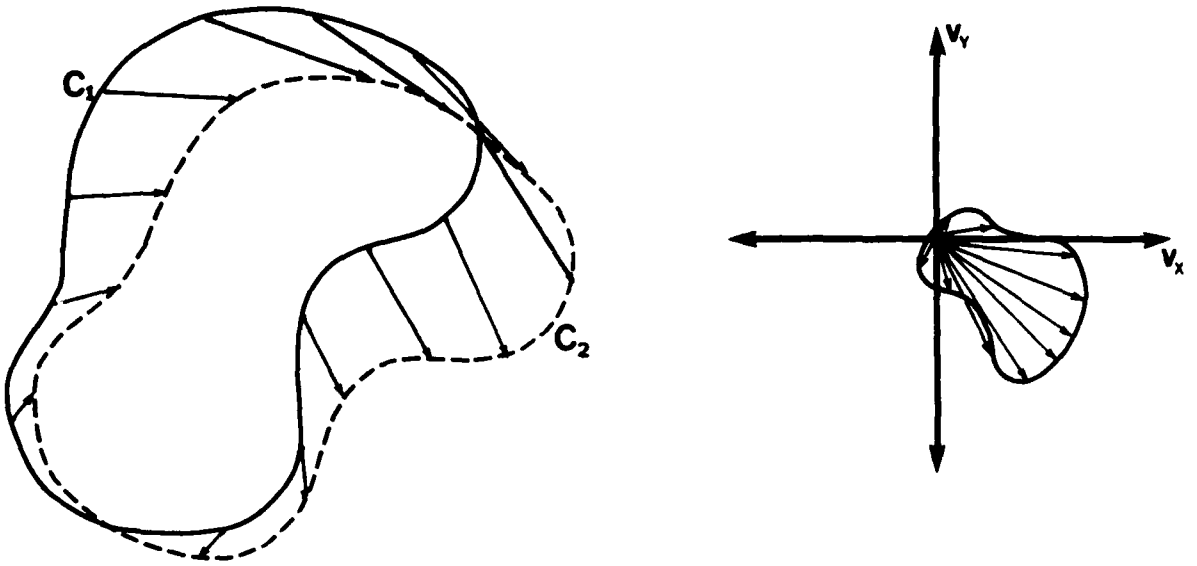
$$\mathbf{V}(s) = M(\mathbf{r} \times \omega\mathbf{n} + \mathbf{d}) = \omega z(s)\begin{bmatrix} n_2 \\ -n_1 \end{bmatrix} + \omega n_3 \begin{bmatrix} -y(s) \\ x(s) \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \tag{2}$$

$M$ denotes the matrix which performs the orthographic projection. The first term in the resulting expression describes the component of the velocity field due to rotation in depth about an axis parallel to the image plane (the axis $\mathbf{n} = [n_1, n_2, 0]^T$); the second term is the component due to motion in the image plane (rotation about the axis $\mathbf{n} = [0, 0, n_3]^T$), and the third term is the translation component.

Consider the restricted case of rigid motion in the image plane; the velocity field now corresponds to the combination of a translation, and rotation about the axis $\mathbf{n} = [0, 0, 1]^T$. Thus, $\mathbf{V}(s)$ is given by:

$$\mathbf{V}(s) = \omega \begin{bmatrix} -y(s) \\ x(s) \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \tag{3}$$

$\mathbf{V}(s)$ is simply a translation, rotation and scaling of the image curve $(x(s), y(s))$, as illustrated in Figure 9. In Figure 9a, the curve $C_1$ undergoes a small rotation and translation in the image plane to yield the curve $C_2$. The arrows indicate local velocity vectors along the curve. In Figure 9b, these velocity vectors have been translated to a common origin in *velocity space*, where the $x$ and $y$ axes represent the $x$ and $y$ components of velocity. The curve in velocity space has the same shape as the image curve $C_1$; its size is proportional to angular velocity $\omega$, and it is rotated 90° with respect to $C_1$ (this relationship is also used in kinematics [69]).




(a) (b)

Figure 9. Rigid motion in the image plane. (a) The velocity field in the image (b) The velocity vectors in velocity space

The additional translation of the curve in the image yields the same translation of the curve in velocity space. In the case of pure translation, the image of the velocity field in velocity space degenerates to a single vector. In general, the explicit use of the velocity space aids in the visualization of properties of the motion of curves, and provides a tool for establishing theoretical properties of the velocity field.

For the simple case of rigid motion in the image plane, this relationship between the shape of the curve and the velocity field is not restricted to continuous motion of the curve. For discrete motion of a curve, we will use the term *displacement field* for the set of vectors which describe the discrete displacement of points on the curve. If we let $\sigma$ be a discrete angular rotation of the curve in the image plane, then the displacement field $V_d(s)$ is given by:

$$V_d(s) = \begin{bmatrix} \cos\sigma - 1 & \sin\sigma \\ -\sin\sigma & \cos\sigma - 1 \end{bmatrix} \begin{bmatrix} x(s) \\ y(s) \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \tag{4}$$

$V_d(s)$ is also given by a scaling and rotation of the projected image curve $(x(s), y(s))$. In this case, the scale factor $k$ is given by:

$$k = \sqrt{(\cos\sigma - 1)^2 + \sin^2\sigma} = \sqrt{2(1 - \cos\sigma)} \tag{5}$$

The angle of rotation $\alpha$ between the image curve $(x(s), y(s))$ and the corresponding curve in velocity space, is given by:

$$\tan\alpha = \frac{\sin\sigma}{\cos\sigma - 1} = -\cot(\frac{\sigma}{2}) \tag{6}$$

For small $\sigma$, $k \approx \sigma$ and $\alpha \approx \pm 90^o$. As before, an additional translation component simply translates the curve in velocity space.

For rigid motion in the image plane, a simple scheme can be used to construct the velocity field. If we know the true direction of velocity for two points on the contour, we can compute the direction of velocity everywhere as follows: (1) construct the lines perpendicular to the direction of velocity at the two known points, (2) compute the intersection of these two lines, (3) from every point $p_i$ along the contour, construct the line to the intersection point; the true direction of velocity is perpendicular to this line. In Figure 10, we derive the direction at $p_2$, given known directions at $p_1$ and $p_3$.

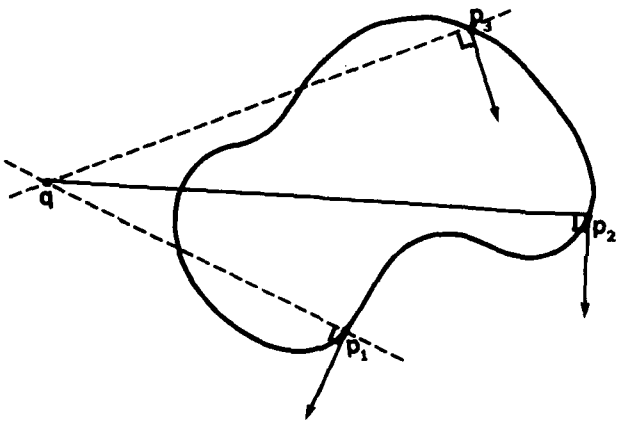



**Figure 10.** Construction of the velocity field for rigid motion in the image plane

This construction is simply locating the point about which the motion can be described as pure rotation. For pure translation, the two lines, from points of known direction of velocity, will be parallel, so the direction of motion everywhere will be equal to the direction of motion of the known points. Certainly, if we knew both the direction and magnitude of velocity at two points along the contour, we could compute the global motion parameters, and hence direction and magnitude of velocity everywhere. However, from the direction of velocity alone at two points on the curve, we can compute the direction of velocity everywhere. If we then know the magnitude of the perpendicular components of velocity along the curve, we can compute both direction and magnitude of velocity along the curve.

There are at least two sources for points of known velocity direction in the image. First, identifiable features, such as terminations, may be tracked in two dimensions. Second, for points at which the perpendicular component of velocity is zero, velocity is constrained to lie along the tangent to the curve. For the case of a smooth, rigid, closed curve moving in the image plane, there must exist at least two points on the curve for which the perpendicular component of velocity is zero. Suppose we focused our velocity field computation at the locations of zero-crossings derived from the image. Since zero-crossing contours are generally closed (except at image boundaries), there will usually be sufficient constraint from the image to solve for the velocity field, in the simple case of rigid motion in the image plane.

### 4.3. The Smoothness Constraint

In this section we will derive a different type of constraint on the velocity field, which will allow us to analyze the projected motion of three-dimensional objects

allowed to move freely in space, and deform over time. The specific analysis will assume that we have measured the perpendicular components of velocity along contours in the image. However, the general constraint that we present may be utilized in other motion measurement schemes as well.

The expression (2) related $\mathbf{V}(s)$ to the global motion parameters $\omega$ and $\mathbf{n}$, and the shape of the curve $C = (x(s), y(s), z(s))$. The relationship between $\mathbf{V}(s)$ and $C$ is quite simple. If we map the projected two-dimensional velocity vectors along the curve to a common origin in velocity space, their endpoints map out a scaled, $90^o$ rotation of the projected image curve $(x(s), y(s))$, with an additional distortion along one direction. This distortion is directed perpendicular to the axis $\mathbf{n} = [n_1, n_2, 0]^T$, and is scaled by the $z$ component of the curve, $z(s)$. This relationship implies, for example, that if we have a smooth curve in motion, it must generate a smoothly varying velocity field. The real world consists predominantly of solid objects, whose surfaces are generally smooth compared with their distance from the viewer. Thus, intuitively, we seek a velocity field which is consistent with the constraints we derive from the changing image, and which varies smoothly. A single solution might be obtained by finding the velocity field which varies as little as possible. A similar argument was used by Horn and Schunck [35] to motivate the use of a smoothness constraint for the optical flow computation. In our case, we seek a smooth velocity field along a contour.

To achieve this, we need some means of measuring the variation in velocity along a contour. There are various ways in which this could be done. For example, we could measure the change in direction of velocity as we trace along the contour. Total variation of the velocity field could then be defined as the total change in direction over the entire contour. A second definition involves measuring the change in magnitude of velocity along the contour. This leads to a velocity field solution for which speed is as uniform as possible along the contour. Finally, we could measure the change in the full velocity vector, $\frac{\partial \mathbf{V}(s)}{\partial s}$, incorporating both the direction and magnitude of velocity.

In order to define the variation of the velocity field more formally, first recall the decomposition of velocity into components tangent and perpendicular to the curve:

$$\mathbf{V}(s) = v^{\top}(s)\mathbf{u}^{\top}(s) + v^{\perp}(s)\mathbf{u}^{\perp}(s) \tag{1}$$

$\mathbf{u}^{\top}(s)$, $\mathbf{u}^{\perp}(s)$ and $v^{\perp}(s)$ can be measured directly from the changing image. $v^{\top}(s)$ is unknown, and must be recovered in order to compute the velocity field $\mathbf{V}(s)$. Aside from knowing $v^{\perp}(s)$ everywhere along the curve, there may also be points at which the direction and magnitude of velocity, and hence both $v^{\perp}(s)$ and $v^{\top}(s)$, are known. In addition, the direction of velocity alone, and hence the ratio $\frac{v^{\perp}(s)}{v^{\top}(s)}$, may be known at points on the curve, for example, where $v^{\perp}(s_i) = 0$ (Section 4.2).

We can now consider a more formal means for measuring the variation in the velocity field. Mathematically, this can be accomplished by defining a functional $\Theta$, which maps the space of all possible vector fields (along the contour), $\mathbf{V}$, into the real numbers: $\Theta$:$\mathbf{V} \mapsto \Re$. This functional should be such that the smaller the variation in the velocity field, the smaller the real number assigned to it. Two candidate velocity fields may then be compared, by comparing their corresponding real numbers. This raises the question of what functional should be used to measure the variation of a velocity field. In the remainder of this section, we will evaluate

a set of possible functionals, based on the three measures of variation that we previously presented informally: (1) variation in $V(s)$, (2) variation in the direction of velocity, and (3) variation in the magnitude of velocity, all with respect to the curve.

(1) Variation in $V(s)$

A scalar measure of the local variation of $V(s)$ with respect to the curve is given by $|\frac{\partial V(s)}{\partial s}|$, shown in Figure 11a. Two nearby velocity vectors along the image curve are translated to a common origin in velocity space, where the vector $\frac{\partial V(s)}{\partial s}$ is shown with a dotted arrow. For convenience of notation, we will omit the argument to

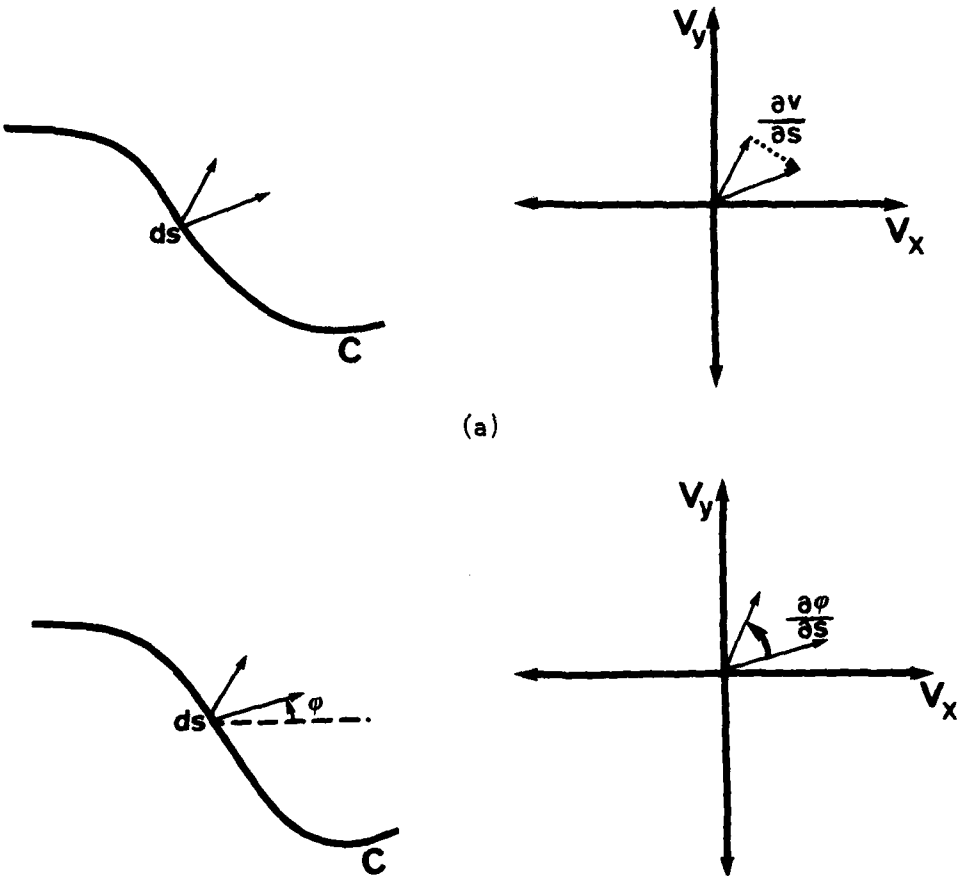



**Figure 11.** Measuring variation in the velocity field (a) Change in the full velocity vector (b) Change in direction of velocity

$V(s)$, writing $|\frac{\partial V}{\partial s}|$. A measure of the total variation of the velocity field along the curve may then be given by the functional:

$$\Theta(V) = \int |\frac{\partial V}{\partial s}| ds$$

We may also consider variations on this functional, involving higher order derivatives, or higher powers, such as:

$$\Theta(\mathbf{V}) = \int |\frac{\partial^2 \mathbf{V}}{\partial s^2}| ds \qquad \text{or} \qquad \Theta(\mathbf{V}) = \int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds$$

(2) Variation in Direction

Let the direction of velocity be given by the angle $\varphi$, measured in the clockwise direction from the horizontal, as shown in Figure 11a. In Figure 11b, $\frac{\partial \varphi}{\partial s}$, for two nearby velocity vectors along the image curve, is shown in velocity space. Total variation of direction along the curve could be given by functionals such as the following:

$$\Theta(\mathbf{V}) = \int |\frac{\partial \varphi}{\partial s}| ds$$

or variations involving higher order derivatives, or higher powers.

(3) Variation in Magnitude

Finally, we could measure the change in magnitude of velocity alone, using functionals such as:

$$\Theta(\mathbf{V}) = \int \frac{\partial |\mathbf{V}|}{\partial s} ds$$

Again, we could also consider variations on this measure.

The functional that we use to measure smoothness may also incorporate a measure of the velocity field itself, rather than strictly utilizing changes in the velocity field along the curve. For example, we could incorporate a term which is a function of $|\mathbf{V}|$. This might be useful if we sought a velocity field which also exhibits the least total motion. In addition, the functional could become arbitrarily complex in its combination of $|\frac{\partial \mathbf{V}}{\partial s}|$, $|\frac{\partial \varphi}{\partial s}|$, $\frac{\partial |\mathbf{V}|}{\partial s}$, or higher order derivatives.

We have at least three means of evaluating these measures of smoothness. From a mathematical point of view, there should exist a unique velocity field which minimizes our particular measure of smoothness; this requirement imposes a set of mathematical constraints on our functional. Second, the velocity field computation should yield physically plausible solutions. Finally, if we suggest that such a smoothness constraint underlies the motion computation in the human visual system, this minimization should yield a velocity field consistent with human motion perception.

An examination of these smoothness measures from a physical and mathematical point of view suggests that a measure involving the full velocity vector, such as $\Theta(\mathbf{V}) = \int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds$, is most appropriate for the velocity field computation [37]. Of particular importance are the mathematical properties of this functional. It can be shown that, given a simple condition on the constraints that we derive from the image, there exists a unique velocity field which satisfies our constraints, and minimizes $\int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds$. This condition is almost always satisfied by our initial motion measurements. To obtain this result, we take advantage of the analysis used by Grimson [70] for evaluating possible functionals for performing surface interpolation from stereo data. The basic mathematical question is, what conditions on the form of the functional, and the structure of the space of velocity fields, are needed to

guarantee the existence of a unique solution? These conditions are captured by the following theorem (see also [71]):

**Theorem**: *Suppose there exists a complete semi-norm $\Theta$ on a space of functions $H$, and that $\Theta$ satisfies the parallelogram law. Then, every nonempty closed convex set $E \subset H$ contains a unique element $v$ of minimal norm, up to an element of the null space. Thus, the family of minimal functions is*

$$\{v + s \mid s \in S\}$$

*where*

$$S = \{v - w \mid w \in E\} \cap \mathcal{N}$$

*and $\mathcal{N}$ is the null space of the functional*

$$\mathcal{N} = \{u \mid \Theta(u) = 0\}.$$

It can be shown that the functional $\{\int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds\}^{\frac{1}{2}}$ is a complete semi-norm, which satisfies the parallelogram law. Second, the space of all possible velocity fields, which satisfy the constraints derived from the image, is convex. It then follows from the above theorem that this space contains a unique element of minimal norm, up to an element of the null space. Since our smoothness measure is non-negative, minimizing $\{\int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds\}^{\frac{1}{2}}$ is equivalent to minimizing $\int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds$.

The null space in this case is the set of constant velocity fields, since $\int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds = 0$ implies $|\frac{\partial \mathbf{V}}{\partial s}| = 0$ everywhere, which implies $\mathrm{V}(s)$ constant. Suppose we have a point $(x(s_i), y(s_i))$ on the curve, where $v^{\perp}(s_i)$ is known. This measurement constrains the velocity $V(s_i)$ to lie along a line parallel to the tangent of the curve at this point, as shown in Figure 12. Suppose we have a velocity field which is

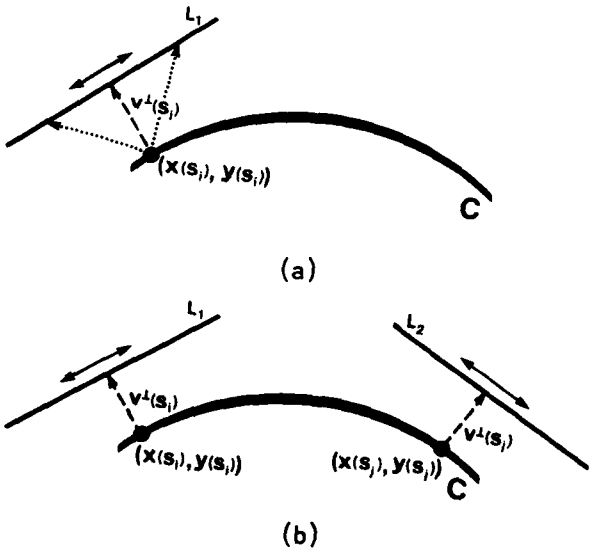



**Figure 12.** Uniqueness of the velocity field. (a) Constraint provided by a single measurement (b) The constraint imposed by two measurements

consistent with this measure. We can now only add a uniform translation component along the direction of this line, and still obtain a velocity field consistent with this local measure. If $v^{\perp}(s)$ is known at a second point $(x(s_i), y(s_i))$, for which the

direction of the tangent is different (see Figure 12b), then we can only add a uniform translation component along this second direction, and still obtain a velocity field consistent with $v^{\perp}(s_j)$. However, we cannot add a uniform translation to the entire velocity field, which is consistent with both local measurements. Thus, we conclude the following: If $v^{\perp}(s)$ is known at two points, for which the orientation of the curve is different, then there exists a unique velocity field which satisfies the known velocity constraints and minimizes $\int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds$. An extended straight line will not yield measurements for two different orientations, but in all other cases, there will be sufficient information along a contour to guarantee a unique solution to the velocity field.

We can apply the constraint of least variation and compute a projected two-dimensional velocity field for any three-dimensional surface, whether rigid or non-rigid, undergoing general motion in space. If we measure the variation in the full velocity vector along a contour in the image, using a functional such as $\int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds$, we are guaranteed that there exists a unique solution to the velocity field computation that minimizes this variation. While it is not yet clear that the general smoothness constraint, or the particular measure $\int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds$, is the most appropriate for the motion computation, it is important that this measure satisfies certain essential mathematical requirements, that the other measures do not. For example, the use of a functional incorporating only a measure of velocity direction, which will attempt to make the local velocity vectors as parallel as possible, does not yield functionals which are semi-norms, and consequently, does not lead to a unique velocity field solution. For a scheme to underly the motion computation in the human visual system, it is essential that it be mathematically well-founded.

We should note that an advantage to applying the smoothness constraint along contours is that the minimization of variation in the velocity field is performed along one-dimension, rather than over two dimensions, as in the case of Horn and Schunck's computation of the optical flow [35]. Secondly, to apply the smoothness constraint over an area of the image, it is necessary to specify a neighborhood size, within which constraints will be combined, and smoothness imposed on the velocity field. Unless we can define surface boundaries prior to the velocity field computation, specifying an appropriate area of the image can be difficult. In general, the extent of contours is more highly correlated with single surfaces. The smoothness constraint can be applied to single contours, reducing the problem of integrating motion measurements across object boundaries. Finally, there exist several standard algorithms for the solution of optimization problems such as this (see, e.g. [37]).

### 4.4. Deriving Additional Constraints from the Image

In the previous section, we used two sources of constraint on the velocity field computation. From the image, we utilized a single curve at a particular moment in time, together with the instantaneous measurements of the perpendicular component of velocity along the curve. As a second source of constraint, we computed the velocity field consisent with these image constraints, which exhibited the least variation along the curve. Additional constraints can be derived from the image if we do not restrict ourselves to the use of instantaneous measurements; for example, we may utilize a second curve, at some time later. If the time interval is small, then the displacement of points along the curve will also be small. We can then require that each point on the first curve project to a point on the second, with a velocity consistent with the instantaneous perpendicular component of velocity $v^{\perp}(s)$; this assumes that velocity is constant over the time interval separating the

two curves. In addition, we could still compute the velocity field which exhibits the least variation.

This approach may yield a simpler, more robust algorithm for the velocity field computation, because it utilizes more constraint from the image. However, it has the disadvantage that we may not be able to obtain the theoretical uniqueness results that were possible when we considered the perpendicular components of velocity as a sole source of constraint. A simple example, in which the velocity field solution is not unique is shown in Figure 13. Suppose we are given the initial constraints shown in Figure 13a. The arrows indicate the perpendicular components of velocity along the first curve, and the dotted line indicates the second curve. There are two velocity field solutions consistent with these constraints, shown in Figures 13b and c, corresponding to the two directions of rotation of the circle. Both velocity fields exhibit the same total variation. In general, theoretical results on uniqueness may be more difficult to obtain for this approach to the velocity field computation. The use of instantaneous motion measurements alone, together with the additional smoothness constraint, as discussed in Section 4.3, would yield the velocity field given by the vectors in Figure 13a, corresponding to pure expansion of the circle. The additional constraint of the second curve leads to a different solution.

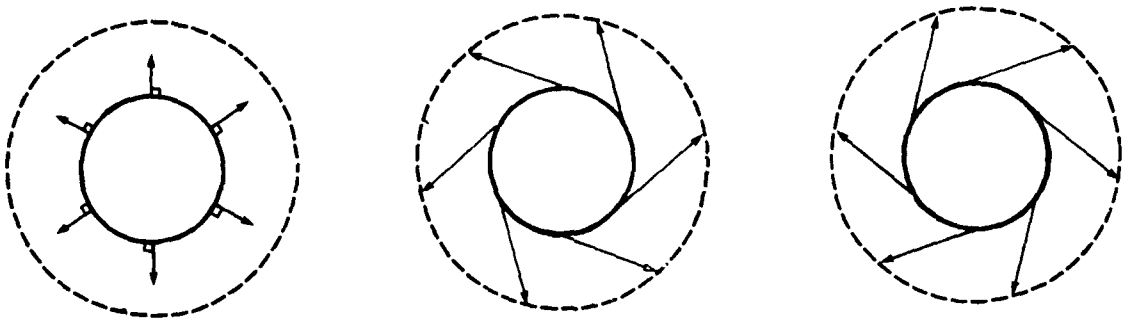



**Figure 13.** Ambiguity of the velocity field. (a) The initial constraints (b) Rotation of the circle to the left (c) Rotation to the right

The availability of the second curve may simplify the velocity field computation in the following way. The perpendicular component of velocity, measured at a point $p$ on the first curve, constrains the velocity vector at $p$ to project to a point along the line $l$ in the second frame, shown in Figure 14. If in addition, $p$ must project to a point $q$ on the second curve, possible candidates for $q$ may be given by the intersection of $l$ with the second curve. In practice, there will be error in the measurement of $v^{\perp}(s)$, and $v^{\perp}(s)$ may not be constant over small time intervals. As a consequence, we should consider a band in the second frame, to which $p$ must project. Candidates for $q$ are then given by the intersection of this band with the second curve, shown in Figure 14. If the curve has fairly high local curvature, or undergoes rotation, then this intersection alone provides considerable constraint on the velocity field. However, in the worst case of an extended line undergoing pure translation, the second curve offers limited additional constraint. The computation of a precise velocity field requires further analysis of constraints derived from the image, together with additional assumptions. We are presently exploring algorithms which utilize the smoothness constraint for this subsequent computation.

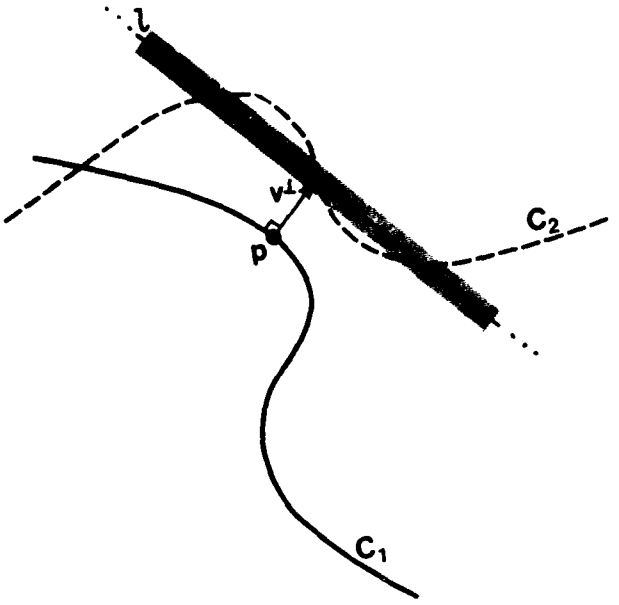



Figure 14. Use of the constraint provided by a second curve

### 4.5. Summary

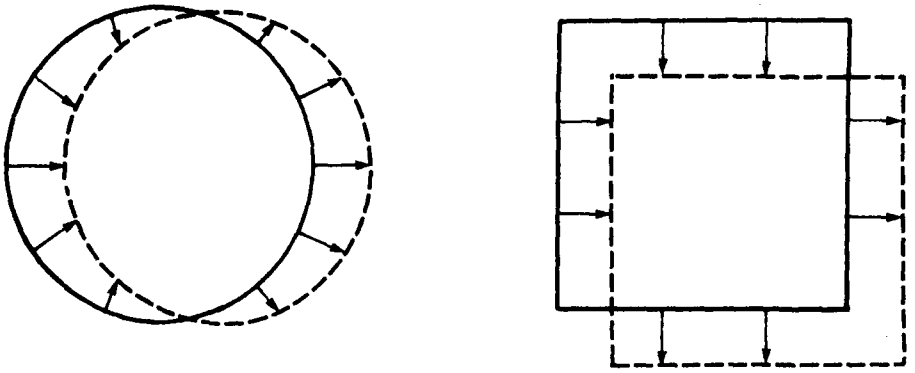
We have considered various additional constraints which may be used in the computation of the velocity field from initial motion measurements derived from the changing image. These constraints range from the restricted assumption of pure translation to the general constraint of smoothness of the velocity field, which allows for the arbitrary movement of rigid or non-rigid objects in space. The use of different constraints results in considerable variation in the classes of motion which may be analyzed, the type of algorithm, and the extent of theoretical analysis required to formulate a well-defined computational problem. In analyzing these constraints, we have so far restricted ourselves to addressing purely computational issues. In the next section, we discuss implications for the biological computation of motion. If the human visual system does in fact compute a detailed velocity field, it is likely to use as much constraint as possible from the changing image, together with the least restrictive additional constraints as necessary, to compute a unique velocity field.

## 5. Some Implications Concerning the Biological Computation of Motion

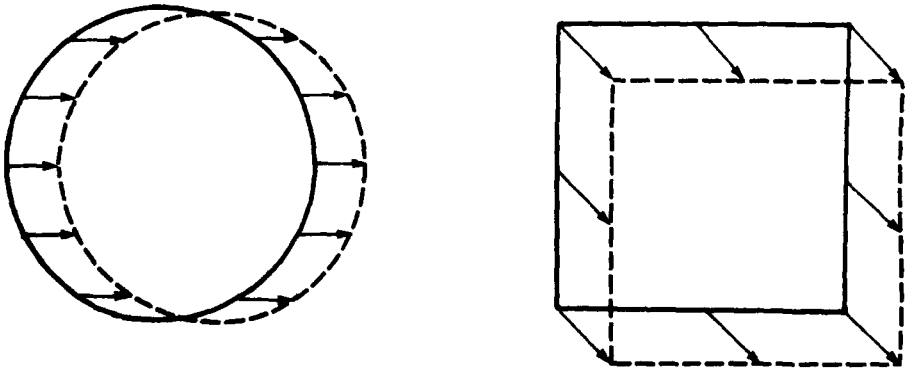
In this section, we summarize the above discussion by presenting a list of the basic proposals that have been made for the computation of motion. In addition, we discuss some of the implications of these proposals for the human visual system.

(i) An underlying assumption of this work is that the local velocity field is explicitly computed and represented. For the human visual system, the idea that there exists an explicit computation of motion, which is different from the description of motion that could be provided by initial motion detectors, can be motivated by simple

examples. In Figure 15, we show a circle and square undergoing pure translation. Initial motion measurements provide the component of motion in the direction perpendicular to the local orientation of intensity changes in the image, shown in Figure 15a. Our perception of the movement of the figures is pure translation, indicated by the set of velocity vectors in Figure 15b. A third example is that of the rotating and translating curve of Figure 9. While it is not clear whether we are capable of explicitly representing the local velocity field around the contour, we do perceive the movement as the rotation and translation of a rigid curve. Such an interpretation is not explicit in the initial motion measurements. For tasks such as the detection of a sudden movement, or separation of objects on the basis of differential motion, a precise local velocity field may not be necessary. However, to compute three-dimensional structure from motion, a more detailed computation of the velocity field, or an explicit correspondence of elements between frames, is required.



(a)



(b)

**Figure 15.** Computing the local velocity field. (a) The initial motion measurements; (b) The velocity field corresponding to translation

(ii) The analysis of motion has been separated into two distinct stages; first, the measurement of motion, and second, the use of motion for tasks such as object

segmentation and structure from motion. This raises the question of whether the interpretation of three-dimensional structure can influence the computation of the two-dimensional velocity field. For example, does the assumption of rigidity, examined in Ullman's work [13], enter into the velocity field computation? Psychological experiments [13] suggest that the long range motion correspondence is not influenced by the intrepreted three-dimensional structure of a single view. The short range process may be similar.

(iii) We support the idea that there exists two processes for analyzing motion, corresponding to Braddick's long range and short range processes. We suggest that the long range process is based on a token-matching scheme, while the short range process is intensity-based. If this view is valid, it raises the following questions. How do the long range and short range processes interact? Do subsequent computational tasks, such as object segmentation or structure from motion utilize the results of one or the other process? The work of Petersik [72] suggests that the long range process may be crucial to the recovery of structure from motion. Finally, it is interesting that neurophysiological studies have revealed many units which are responsive to, or selective for stimuli undergoing continuous motion. Little is known about the long range process at the neurophysiological level. One obvious question is, where in the visual system can apparent motion phenomena be observed in the response of single units? Motion sensitive units (for example in areas V1 and STS or MT of the monkey) could be tested for apparent motion response by flashing bars at stationary locations using relatively wide separations (that is, wider than the largest size of simple cells at the tested eccentricity). If long range motion units can be identified, it may become possible to go a step further and investigate the relationship between the psychophysically established correspondence rules and their neurophysiological correlates.

(iv) We suggest that the initial stage of motion analysis consists of the measurement of the perpendicular component of velocity along zero-crossing contours. This can be examined through neurophysiological and psychological studies. In regard to neurophysiology, are the class of directionally-selective simple cells detecting the motion of zero-crossings in their input from the LGN? This is now under investigation [Richter, personal communication]; initial results tend to support this claim. Psychophysical experiments can test whether perceived motion is consistent with the motion of zero-crossings.

(v) We propose that the local motion measurements are then integrated along zero-crossing contours. Again, this may be explored through both neurophysiology and psychology. If the motion integration problem is fundamental to motion analysis, one may expect to find neural mechanisms within the visual system that are involved in this task. Most of the motion sensitive units studied so far do not seem suitable for the integration stage. Motion selective cells in the primary visual cortex of the cat and the monkey respond primarily to edges and bars. To activate such a unit the stimulus must have the preferred orientation, and move in the preferred direction. In contrast, promising candidates for the integration phase would dissociate the effects of orientation and direction of movement, ideally exhibiting specificity for direction of motion but not for orientation. Furthermore, the direction specificity of such a unit is expected to depend on the range of orientations spanned by the stimulus. There are indications for the possible existence of such units in the posterior bank of the superior temporal sulcus of the rhesus monkey [73]. For psychophysical experimentation, there are at least two questions; first, is the motion that we perceive forced to be consistent at least with the sign of the local motion measurements along zero-crossing contours, or can it be overridden, for example, by the long range process, or by the history of the motion? Second, if the integration does take place, are measurements combined over neighborhoods in the image, or

along contours? Wallach's [74] demonstrations suggest that the integration may take place along contours.

(vi) Additional assumptions are required for the motion integration problem. Regarding the human system, we may first ask what constraints are derived from the changing image. Does the human visual system strictly utilize instantaneous measurements of velocity, or is a second curve, at some small time interval later, also used to constrain the velocity field. Do we utilize an additional constraint on smoothness of the velocity field, as described here? A constraint such as smoothness may be the least restrictive constraint which allows objects to move freely in space, and deform, but which still allows for the computation of a unique velocity field. Psychophysical experimentation is necessary to determine whether the velocity field that we perceive is the smoothest one possible. Both the short and long range processes face the fact that in general, the motion of elements is not specified uniquely by information in the changing image; do the additional assumptions governing the computation of velocity or correspondence differ in the two processes, or do they differ only in the constraints that are utilized from the changing image?

(vii) Finally, the motion measurement problem has some implications for the interpretation of structure from motion. It has been shown [17] that three-dimensional shape can be recovered locally, from the instantaneous velocity field. The interpretation is sensitive, however, to small errors in the measured velocity. In light of the inherent difficulties in measuring the velocity field precisely, recovery methods that rely solely on the instantaneous velocity field appear unlikely. For the reliable recovery of three-dimensional structure from motion, processes that integrate motion over time are probably required.

## References


1. J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, W. H. Pitts: Proc. IRE. **47**, 1940-1951 (1959)
2. W. Reichardt, T. Poggio: Biol. Cyb. **35**, 81-100 (1979)
3. H. R. Maturana, S. Frenk: Science **142**, 977-979 (1963)
4. H. B. Barlow, R. N. Hill: Science **139**, 412-414 (1963)
5. O. Grusser, U. Grusser-Cornehls: "Neuronal Mechanisms of Visual Movement Perception and Some Psychophysical and Behavioral Correlation," *Handbook of Sensory Physiology*, Vol. VII/3A, ed. by R. Jung, (Springer-Verlag, Berlin, 1973)
6. D. H. Hubel, T. Wiesel: J. Physiol. (Lond.) **160**, 106-154 (1962)
7. D. H. Hubel, T. Wiesel: J. Physiol. (Lond.) **195**, 215-243 (1968)
8. D. N. Lee, P. E. Reddish: Nature **293**, 293-294 (1981)
9. H. Wallach, D. N. O'Connell: J. Exp. Psych. **45**, 205-217 (1953)
10. G. Johansson: Perception and Psych. **14**, 201-211 (1973)
11. G. Johansson: Sci. Am. **232**, 76-88 (1975)
12. D. Marr, S. Ullman: Proc. Roy. Soc. Lond. B **211**, 151-180 (1981)
13. S. Ullman: *The Interpretation of Visual Motion* (MIT Press, Cambridge and London, 1979)
14. S. Ullman: Proc. Roy. Soc. Lond. B **203**, 405-426 (1980)
15. K. Prazdny: Biol. Cyb. **36**, 87-102 (1980)

16. W. F. Clocksin: Perception **9**, 253-269 (1980)
17. H. C. Longuet-Higgins, K. Prazdny: Proc. Roy. Soc. Lond. B **208**, 385-397 (1981)
18. E. A. Smith, D. R. Phillips: IEEE Trans. Comp. **c-21**, 715-729 (1972)
19. J. A. Leese, C. S. Novak, V. R. Taylor: Patt. Recog. **2**, 279-292 (1970)
20. R. L. Lillestrand: IEEE Trans. Comp. **c-21**, 654-659 (1972)
21. K. Wolferts: Proc. Int. Joint Conf. Patt. Recog. **2** 1-2 (1974)
22. H. H. Bell, J. S. Lappin: Perception and Psych. **11**, 263-268 (1973)
23. A. J. Pantle, L. Picciano: Science **193**, 500-502 (1976)
24. J. T. Petersik, K. I. Hicks, A. J. Pantle: Perception **7**, 371-383 (1978)
25. R. Jain, W. N. Martin, J. K. Aggarwal: Comp. Graph. Image Proc. **11**, 13-34 (1979)
26. R. Jain, D. Militzer, H. H. Nagel: Proc. IJCAI 5 612-618 (1977)
27. R. Jain, H. H. Nagel: IEEE Trans. PAMI **PAMI-1**, 206-214 (1979)
28. H. H. Nagel: Comp. Graph. Image Proc. **7**, 149-194 (1978)
29. B. Hassenstein, W. Reichardt, W.: Chlorophanus. Z. Naturf. **IIb**, 513-524 (1956)
30. H. B. Barlow, R. W. Levick: J. Physiol. (Lond.) **173**, 477-504 (1965)
31. R. C. Emerson, G. L. Gerstein: J. Neurophysiol. **40**, 136-155 (1977)
32. T. Poggio, W. Reichardt: Quart. Rev. Biophys. **9**, 377-438 (1976)
33. V. Torre, T. Poggio: Proc. Roy. Soc. Lond. B **202**, 409-416 (1978)
34. C. I. Fennema, W. B. Thompson: Comp. Graph. Image Proc. **9**, 301-315 (1979)
35. B. K. P. Horn, B. G. Schunck: Artif. Intel. **17** 185-203 (1981)
36. L. S. Davis, Z. Wu, and H. Sun: In *Proc. ARPA Image Understanding Workshop*, ed. by L. S. Baumann (Science Applications Inc., Arlington, VA, 1979)
37. E. Hildreth: In *Proc. Workshop on Rep. and Control.* (IEEE Computer Society Press, Los Angeles, 1982)
38. J. P. Frisby: Vis. Res. **12**, 1145-1166 (1972)
39. P. A. Kolers: *Aspects of Motion Perception* (Pergamon Press, New York, 1972)
40. A. J. Pantle: J. Opt. Soc. Am. **63**, 1280 (1973)
41. J. Ternus: Psych. Forsch. **7**, 81-136 (1926) Translated in: *A Source Book of Gestalt Psychology*, ed. by W. D. Ellis, (Humanities Press, New York, 1967)
42. D. Navon: J. Exp. Psych. (Hum. Perc. Perf.) **2**, 130-138 (1976)
43. S. Ullman: Perception **7**, 683-693 (1978)
44. S. Ullman: Perception **9**, 617-626 (1981)
45. J. L. Potter: IEEE Trans. Sys., Man, Cyb. **SMC-5**, 390-394 (1975)
46. J. L. Potter: Comp. Graph. Image Proc. **6**, 558-581 (1977)
47. J. K. Aggarwal, R. O. Duda: IEEE Trans. Comp. **c-24**, 966-976 (1976)
48. W. K. Chow, J. K. Aggarwal: IEEE Trans. Comp. **c-26**, 179-185 (1977)
49. J. W. Roach, J. K. Aggarwal: IEEE Trans. PAMI **PAMI-1**, 127-135 (1979)

50. D. Lawton: Comp. Graph. Image Proc. *in press* (1982)
51. D. Marr: Phil. Trans. Roy. Soc. Lond. B **275**, 483-524 (1976)
52. D. Marr: *VISION* (W. H. Freeman Co., San Francisco, 1981)
53. M. Wertheimer: Zeitchrift fur Psych. **61**, 161-265 (1912)
54. W. Neuhaus: Archiv fur die Gesamte Psych. **75**, 315-348 (1930)
55. S. M. Anstis, B. J. Rogers: Vis. Res. **15**, 957-961 (1975)
56. S. M. Anstis: Phil. Trans. Roy. Soc. Lond. B. **290**, 153-168 (1980)
57. O. J. Braddick: Vis. Res. **13**, 355-369 (1973)
58. O. J. Braddick: Vis. Res. **14**, 519-527 (1974)
59. O. J. Braddick: Phil. Trans. Roy. Soc. Lond. B. **290**, 137-151 (1980)
60. J. T. Petersik, K. I. Hicks, K. I., A. J. Pantle: Perception **7**, 371-383 (1978)
61. M. Riley: *The Representation of Image Texture* (MIT AI TR-649, 1981)
62. E. C. Hildreth: Comp. Graph. Image Proc. *in press* (1982)
63. D. Marr, E. C. Hildreth: Proc. Roy. Soc. Lond. B. **207**, 187-217 (1980)
64. D. Marr, T. Poggio: Proc. Roy. Soc. Lond. B. **204**, 301-328 (1979)
65. J. Richter, S. Ullman: Biol. Cyb. **43**, 127-145 (1982)
66. J. O. Limb, J. A. Murphy: Comp. Graph. Image Proc. **4**, 311-327 (1975)
67. J. Batali, S. Ullman: In *Proc. ARPA Image Understanding Workshop*, ed. by L. S. Baumann (Science Applications Inc., Arlington, VA, 1979)
68. A. Bruss, B. K. P. Horn: *Passive Navigation* (MIT AI Memo 661, 1981)
69. R. Hartenberg, J. Denauit: *Kinematic Synthesis of Linkages* (McGraw-Hill Co., New York, 1964)
70. W. E. L. Grimson: *From Images to Surfaces. A Computational Study of the Human Early Visual System* (MIT Press, Cambridge and London, 1981)
71. W. Rudin: *Functional Analysis*, (McGraw-Hill Co., New York, 1973)
72. J. T. Petersik: Perception **9**, 271-283 (1980)
73. S. M. Zeki: J. Physiol. **236**, 549-573 (1974)
74. H. Wallach: "On Perceived Identity: 1. The Direction of Motion of Straight Lines," In *On Perception*, ed. by H. Wallach, (Quadrangle, N.Y., 1976)